

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
7 December 2000 (07.12.2000)

PCT

(10) International Publication Number  
**WO 00/72869 A1**

- (51) International Patent Classification<sup>7</sup>: A61K 38/16, 38/03, C07K 14/00
- (74) Agent: ELBING, Karen, L.; Clark & Elbing, LLP, 176 Federal Street, Boston, MA 02110-2214 (US).
- (21) International Application Number: PCT/US00/15077
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (22) International Filing Date: 1 June 2000 (01.06.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/137,032 1 June 1999 (01.06.1999) US
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant: PHYLOS, INC. [US/US]; 128 Spring Street, Lexington, MA 02421 (US).
- Published:**  
— With international search report.
- (72) Inventors: LOHSE, Peter; 50 Golden Ball Road, Weston, MA 02493 (US). WRIGHT, Martin, C.; 812 Memorial Drive #1105, Cambridge, MA 02139 (US). MCPHERSON, Michael; 142 Butler Avenue, Providence, RI 02906 (US).
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 00/72869 A1

(54) Title: METHODS FOR PRODUCING 5'-NUCLEIC ACID-PROTEIN CONJUGATES

(57) Abstract: Disclosed herein is a method for generating a 5'-nucleic acid-protein conjugate, the method involving: (a) providing a nucleic acid which carries a reactive group at its 5' end; (b) providing a non-derivatized protein; and (c) contacting the nucleic acid and the protein under conditions which allow the reactive group to react with the N-terminus of the protein, thereby forming a 5'-nucleic acid-protein conjugate. Also disclosed herein are 5'-nucleic acid-protein conjugates and methods for their use.

METHODS FOR  
PRODUCING 5'-NUCLEIC ACID-PROTEIN CONJUGATES

5

Background of the Invention

In general, the present invention features methods for the preparation of nucleic acid-protein conjugates.

Nucleic acid-protein conjugates, sometimes referred to as nucleic acid-protein fusions, nucleoproteins or nucleopeptides, are naturally-occurring  
10 bioconjugates which play a key role in important biological processes. In one particular example, such conjugates play a central role in the process of nucleoprotein-primed viral replication (Salas, Ann. Rev. Biochem. 60, 39-71 (1991)). Accordingly, nucleoproteins as well as nucleopeptides may serve as powerful tools for the study of biological phenomena, and may also provide a  
15 basis for the development of antiviral agents.

In addition, conjugates of peptides and nucleic acids have found use in several other applications, such as non-radioactive labels (Haralambidis et al., Nucleic Acids Res. 18, 501-505 (1990)) and PCR primers (Tong et al., J. Org. Chem. 58, 2223-2231 (1993)), as well as reagents in encoded  
20 combinatorial chemistry techniques (Nielsen et al., J.A.C.S. 115, 9812-9813 (1993)). In yet other applications, peptides predicted to have favorable interactions with cell membranes, such as polylysine (Leonetti et al., Bioconjugate Chem. 1, 149-153 (1990)), other highly basic peptides (Vives & Lebleu, Tetrahedron Lett. 38, 1183-1186 (1997)), hydrophobic peptides (Juby et al., Tetrahedron Lett. 32, 879-882 (1991)), viral fusion peptides (Soukchareun et al., Bioconjugate Chem. 6, 43-53 (1995)) and peptide signal sequences (Arar et al., Bioconjugate Chem. 6, 573-577 (1995)), have been coupled to  
25

oligonucleotides to enhance cellular uptake. Peptides able to chelate metals have also been appended to oligonucleotides to generate specific nucleic acid cleaving reagents (Truffert et al., Tetrahedron 52, 3005-3016 (1996)). And peptides linked to the 3'-end of oligonucleotides have been reported to provide  
5 important resistance to 3'-exonucleases (Juby et al., Tetrahedron Lett. 32, 879-882 (1991)).

One particular type of nucleic acid-protein conjugate, referred to as an RNA-protein fusion (Szostak and Roberts, U.S.S.N. 09/007,005; and Roberts and Szostak, Proc. Natl. Acad. Sci. USA 94, 12297-12302 (1997)), has  
10 been used in methods for isolating proteins with desired properties from pools of proteins. To create such fusions, an RNA and the peptide or protein that it encodes are joined during in vitro translation using synthetic RNA that carries a peptidyl acceptor, such as puromycin, at its 3'-end. In this process, the synthetic RNA, which is devoid of stop codons, is typically synthesized by in  
15 vitro transcription from a DNA template followed by 3'-ligation to a DNA linker carrying puromycin. The DNA template sequence causes the ribosome to pause at the 3'-end of the open reading frame, providing additional time for the puromycin to accept the nascent peptide chain and resulting in the production of the RNA-protein fusion molecule.

## 20 Summary of the Invention

The present invention features chemical ligation methods for producing nucleic acid-protein conjugates in good yields. Two different approaches are described. In the first, fusions are formed by a reaction between an unprotected protein carrying an N-terminal cysteine and a nucleic acid  
25 carrying a 1,2-aminothiol reactive group. In the second approach, fusion formation occurs as the result of a bisarsenical-tetracysteine interaction.

Accordingly, in a first aspect, the invention features a method for generating a 5'-nucleic acid-protein conjugate, the method involving: (a) providing a nucleic acid which carries a reactive group at its 5' end; (b) providing a non-derivatized protein; and (c) contacting the nucleic acid and the protein under conditions which allow the reactive group to react with the N-terminus of the protein, thereby forming a 5'-nucleic acid-protein conjugate.

In a related aspect, the invention features a 5'-nucleic acid-protein conjugate which includes a nucleic acid bound through its 5'-terminus or a 5'-terminal reactive group to the N-terminus of a non-derivatized protein.

In various preferred embodiments of these aspects, the nucleic acid is greater than about 20 nucleotides in length; the nucleic acid is greater than about 120 nucleotides in length; the nucleic acid is between about 2-1000 nucleotides in length; the protein is greater than about 20 amino acids in length; the protein is greater than about 40 amino acids in length; the protein is between about 2-300 amino acids in length; the contacting step is carried out in a physiological buffer; the contacting step is carried out using a nucleic acid and a protein, both of which are present at a concentration of less than about 1 mM; the nucleic acid is DNA or RNA (for example, mRNA); the nucleic acid includes the coding sequence for the protein; the N-terminus of the non-derivatized protein is a cysteine residue; the N-terminal cysteine is exposed by protein cleavage; the reactive group is an aminothiols reactive group; the protein includes an  $\alpha$ -helical tetracysteine motif located proximal to its N-terminus; the  $\alpha$ -helical tetracysteine motif includes the sequence cys-cys-X-X-cys-cys, wherein X is any amino acid; the reactive group is a bisarsenical derivative; the conjugate is immobilized on a solid support (for example, a bead or chip); and the conjugate is one of an array immobilized on a solid support.

In another related aspect, the invention features a method for the selection of a desired nucleic acid or a desired protein, the method involving:

(a) providing a population of 5'-nucleic acid-protein conjugates, each including a nucleic acid bound through its 5'-terminus or a 5'-terminal reactive group to the N-terminus of a non-derivatized protein; (b) contacting the population of 5'-nucleic acid-protein conjugates with a binding partner specific for either the nucleic acid or the protein portion of the desired nucleic acid or desired protein under conditions which allow for the formation of a binding partner-candidate conjugate complex; and (c) substantially separating the binding partner-candidate conjugate complex from unbound members of the population, thereby selecting the desired nucleic acid or the desired protein.

In yet another related aspect, the invention features a method for detecting an interaction between a protein and a compound, the method involving: (a) providing a solid support that includes an array of immobilized 5'-nucleic acid-protein conjugates, each conjugate including a nucleic acid bound through its 5'-terminus or a 5'-terminal reactive group to the N-terminus of a non-derivatized protein; (b) contacting the solid support with a candidate compound under conditions which allow an interaction between the protein portion of the conjugate and the compound; and (c) analyzing the solid support for the presence of the compound as an indication of an interaction between the protein and the compound.

In various preferred embodiments of these methods, the method further involves repeating steps (b) and (c); the compound is a protein; the compound is a therapeutic; the nucleic acid is greater than about 20 nucleotides in length; the nucleic acid is greater than about 120 nucleotides in length; the nucleic acid is between about 2-1000 nucleotides in length; the protein is greater than about 20 amino acids in length; the protein is greater than about 40

amino acids in length; the protein is between about 2-300 amino acids in length; the nucleic acid is DNA or RNA (for example, mRNA); the nucleic acid includes the coding sequence for the protein; the N-terminus of the non-derivatized protein is a cysteine residue; the reactive group is an aminothiols  
5 reactive group; the protein includes an  $\alpha$ -helical tetracysteine motif located proximal to its N-terminus; the  $\alpha$ -helical tetracysteine motif includes the sequence cys-cys-X-X-cys-cys, wherein X is any amino acid; the reactive group is a bisarsenical derivative; the conjugate is immobilized on a solid support (for example, a bead or chip); and the conjugate is one of an array immobilized on a  
10 solid support.

As used herein, by a "5'-nucleic acid-protein conjugate" is meant a nucleic acid which is covalently bound to a protein through the nucleic acid's 5' terminus.

By a "nucleic acid" is meant any two or more covalently bonded  
15 nucleotides or nucleotide analogs or derivatives. As used herein, this term includes, without limitation, DNA, RNA, and PNA.

By a "protein" is meant any two or more amino acids, or amino acid analogs or derivatives, joined by peptide or peptoid bond(s), regardless of length or post-translational modification. As used herein, this term includes,  
20 without limitation, proteins, peptides, and polypeptides.

By "derivatize" is meant adding a non-naturally-occurring chemical functional group to a protein following the protein's translation or chemical synthesis. "Non-derivatized" proteins are not treated in this manner and do not carry such non-naturally-occurring chemical functional groups.

25 By a "physiological buffer" is meant a solution that mimics the conditions in a cell. Typically, such a buffer is at about pH 7 and may be at a temperature of about 37°C.

By a "solid support" is meant any solid surface including, without limitation, any chip (for example, silica-based, glass, or gold chip), glass slide, membrane, bead, solid particle (for example, agarose, sepharose, or magnetic bead), column (or column material), test tube, or microtiter dish.

5 By an "array" is meant a fixed pattern of immobilized objects on a solid surface or membrane. As used herein, the array is made up of nucleic acid-protein fusion molecules (for example, RNA-protein fusion molecules). The array preferably includes at least  $10^2$ , more preferably at least  $10^3$ , and most preferably at least  $10^4$  different fusions, and these fusions are preferably arrayed  
10 on a 125 x 80 mm, and more preferably on a 10 x 10 mm, surface.

By a "population" is meant more than one molecule.

The present invention provides a number of advantages. For example, although conjugates of between 2-1000 nucleotides and 2-300 amino acids are preferred, nucleic acid-protein conjugates of any desired molecular  
15 weight may be generated using the methods of the invention because the nucleic acid as well as the protein may be produced independently using well-known synthetic and biological methods. These post-synthetic ligation methods are therefore advantageous over fully synthetic techniques where stepwise buildup of nucleic acid-peptide conjugates generally allows  
20 preparation of only limited size conjugates, typically of less than 20 nucleotides and less than 20 amino acids in length.

In addition, the reactions described herein (for example, the reaction between the N-terminal cysteine and the 1,2-aminothiol reactive group on the nucleic acid) are chemoselective over other nucleophilic groups on the protein,  
25 thus leading to regiospecific links between proteins and nucleic acids. This contrasts with known methods for the synthesis of protein-nucleic acid conjugates which often rely on reactions between a nucleophilic group on the

protein and an electrophile on the nucleic acid moiety (Bayard et al.,  
Biochemistry 25, 3730-3736 (1986); Cremer et al., J. Prot. Chem. 11(5),  
553-560 (1992)). In these reactions, multiple nucleophilic side chains on the  
protein compete for reaction with the electrophile leading to non-specific links  
5 between protein and nucleic acid and thus generating a heterogenous mixture of  
conjugate products.

In yet other advantages, the present ligation reactions work  
efficiently under mild conditions in physiological buffers. Consequently,  
protein structure is not disrupted under the ligation conditions used, and  
10 conjugates carrying functional proteins can be formed. In addition, the present  
ligation reactions work efficiently with reactand concentrations in the  $\mu$ M  
range. Consequently, dilute preparations of protein and nucleic acid can be used  
for conjugate preparation.

The present techniques also provide advantages with respect to the  
15 conjugates themselves. Most notably, the conjugate nucleic acid (for example,  
RNA) is linked to the amino-terminus of the conjugate protein. This type of  
fusion leaves the protein's carboxy-terminus unmodified and is particularly  
beneficial when the carboxy-terminal amino acids are involved with protein  
structure or function, or participate in interactions with other species.

20 In addition, with respect to RNA-protein fusions, efficient ligation in  
aqueous buffers at low concentrations of reactands allows the fusion of nascent  
proteins to their encoding RNAs while bound to the ribosome. Pretranslational  
3'-modification of the mRNA as described for 3'-fusions (Szostak and Roberts,  
U.S.S.N. 09/007,005; and Roberts and Szostak, Proc. Natl. Acad. Sci. USA 94,  
25 12297-12302 (1997)) is unnecessary, because the 3'-end of the mRNA is not  
involved in ligation. Moreover, because of the lack of involvement of the  
3'-end of the RNA in ligation, the present technique facilitates the production of

RNA-protein fusions using RNAs from a variety of sources. In one particular example, RNA (for example, mRNA) libraries with heterogeneous 3'-termini may be readily used for the synthesis of 5'-mRNA-protein fusions. In another example, cellular RNA may be used for fusion formation.

5           Finally, the present invention provides a quantitative advantage for the production of RNA-protein fusions by simplifying ribosome turnover and thereby optimizing fusion synthesis. In particular, because conjugate proteins are linked through their N-termini to conjugate nucleic acids, the fusion products are released in unhindered fashion from the native ribosome following  
10 translation, allowing free ribosomes to undergo further rounds of translation. This multiple turnover allows for the synthesis of larger pools of RNA-protein fusions than is currently available with single turnover at the ribosome (Szostak and Roberts, U.S.S.N. 09/007,005; and Roberts and Szostak, Proc. Natl. Acad. Sci. USA 94, 12297-12302 (1997)).

15           The nucleic acid-protein fusions (for example, the mRNA-protein fusions) of the invention may be used in any selection or in vitro evolution technique. For example, these fusions may be used in methods for the improvement of existing proteins or the evolution of proteins with novel structures or functions, particularly in the areas of therapeutic, diagnostic, and  
20 research products. In addition, 5'-RNA-protein fusions find use in the functional genomics field; in particular, these fusions (for example, cellular mRNA-protein fusions) may be used to detect protein-protein interactions in a variety of formats, including presentation of fusion arrays on solid supports (for example, beads or microchips).

25           Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

### Brief Description of the Drawings

FIGURE 1 is a diagram which illustrates the general approach of the invention for generating nucleic acid-protein conjugates.

FIGURE 2 is a diagram which illustrates the general approach for  
5 generating fusions between a protein and its encoding mRNA on the ribosome.

FIGURE 3 is a diagram which illustrates the 1,2-aminothiol reactive group modifier, "phenyl- $\alpha$ -bromothioacetate."

FIGURE 4 is a diagram which illustrates alkylation of 5'-GMPS-modified RNA with phenyl- $\alpha$ -bromothioacetate.

10 FIGURE 5 is a diagram which illustrates an orthogonal ligation reaction between a nucleic acid carrying a thioester functional group and a protein carrying an N-terminal cysteine.

FIGURE 6 is a diagram which illustrates the formation of nucleic acid-protein conjugates using a bisarsenical-tetracysteine interaction.

15 FIGURE 7 is a diagram which illustrates an exemplary synthetic scheme for the synthesis of a bisarsenical derivative.

FIGURE 8 is a diagram which illustrates a second exemplary synthetic scheme for the synthesis of a bisarsenical derivative.

### Detailed Description

20 The present methods for the synthesis of nucleic acid-protein conjugates are based on chemical ligation reactions which take place between the nucleic acid and the protein components.

In the first approach, the ligation reaction takes place between an unprotected protein carrying an N-terminal cysteine and a nucleic acid carrying  
25 a 1,2-aminothiol reactive group. The ligation reaction is performed generally as described for the synthesis of proteins from protein fragments (see, for

example, Brenner, in Peptides, Proceedings of the Eighth European Peptide Symposium, Beyermann, ed. (North-Holland, Amsterdam, 1967), pp. 1-7; Kemp & Carey, J. Org. Chem. 58, 2216 (1993); Liu & Tam, J. Am. Chem. Soc. 116, 4149 (1994); Dawson et al., Science 266, 776 (1994)). A fast  
5 chemoselective reaction followed by intramolecular amide bond formation leads to a covalent link between the nucleic acid and protein. This reaction requires the protein to carry an N-terminal cysteine and the nucleic acid to carry a 1,2-aminothiol reactive group. The general approach is illustrated in Figure 1. Ligation of a protein to its encoding RNA while bound to the ribosome is  
10 illustrated in Figure 2.

#### Preparation of Proteins for Orthogonal Ligation

The first ligation scheme according to the invention requires the protein to carry an N-terminal cysteine. Such proteins may be easily prepared synthetically using standard chemical synthetic methods. Alternatively,  
15 proteins may be prepared by biological or recombinant methods. These proteins, however, typically do not carry an N-terminal cysteine, instead beginning with an N-terminal methionine residue due to translational initiation at an AUG start codon. Various methods may be utilized to expose a cysteine at the N-terminus of the conjugate protein. In one particular example,  
20 endogenous aminopeptidase activity present in a cellular lysate may be used to remove the N-terminal methionine, thereby exposing the penultimate amino acid at the N-terminus (Moerschell et al., J. Biol. Chem. 265, 19638-19643 (1990)). Alternatively, an N-terminal fragment may be cleaved from each protein in a population of proteins having homogeneous N-termini using a  
25 sequence-specific protease. This cleavage reaction produces a population of proteins, each having an N-terminal cysteine (that is, the amino acid C-terminal

to the cleavage site). Suitable proteases for this purpose include, without limitation, Factor Xa and Enterokinase (both of which are available from New England Biolabs, Inc., Beverly, MA). These proteases are used in accordance with the manufacturer's instructions.

## 5    Preparation of Nucleic Acids for Orthogonal Ligation

The first ligation method of the invention also requires a nucleic acid which carries a 1,2-aminothiol reactive group. This group may be introduced during the synthesis of the nucleic acid or after synthesis (post-synthetically) by means of a 1,2-aminothiol reactive modifier.

10        Nucleic acids or nucleic acid analogs may be synthesized by standard chemical or enzymatic methods. Heterogenous mixtures of nucleic acids (for example, pools of random sequences or cellular mRNA libraries) may also be readily utilized. Preferably, for fusion formation on a ribosome, the RNA utilized contains no inadvertent stop codons.

15        For the incorporation of the thiol or thiophosphate group into the nucleic acid, any of a number of standard techniques may be exploited. For example, thiol groups may be incorporated into DNA by chemical means (see thiolmodifiers, Glen Research, Sterling, Virginia; Raines & Gottlieb, RNA 4, 340-345 (1998); Gundlach et al., Tetrahedron Lett. 38, 4039 (1997); Coleman  
20    & Siedlecki, J. Am. Chem. Soc. 114, 9229 (1992)). Alternatively, terminal thiophosphate groups may be prepared by chemical phosphorylation followed by oxidation with a sulfurizing reagent (Glen Research, Sterling, Virginia).

In yet another approach, thiol and thiophosphate groups may be incorporated into RNA by enzymatic means. In one preferred method for the  
25    generation of 5'-modified RNA, transcription is carried out in the presence of GMPαS, GDPβS or GTPγS, followed by chemical modification of the

5'-thiophosphate group as described, for example, in Burgin & Pace, EMBO Journal 9, 4111-4118 (1990); and Logsdon et al., Anal. Biochem. 205, 36-41 (1992). Alternatively, guanosine derivatives carrying the 1,2-aminothiol reactive group may be used to initiate transcription as described, for example, in Martin & Coleman, Biochemistry 28, 2760-2762 (1989); and Logsdon et al., Anal. Biochem. 205, 36-41 (1992). For any of these techniques, GMP $\alpha$ S may be purchased from Amersham, Buckinghamshire, UK, and GTP $\gamma$ S may be purchased from Fluka, Milwaukee, WI.

A preferred 1,2-aminothiol reactive modifier is

- 10 phenyl- $\alpha$ -bromothioacetate, shown in Figure 3. This compound may be synthesized using the procedure of Gennari et al., Tetrahedron 53(16), 5909-5924 (1997)). Specifically, this compound was prepared as follows. To a cooled (0°C) solution of benzenethiol (0.551 g, 5 mmol, 0.51 ml) in dry dichloromethane (10ml) was added dry pyridine (0.435 g, 5.5 mmol, 0.45 ml).  
15 Bromoacetyl chloride (Fluka, 0.787 g, 5 mmol, 0.417 ml) in dry dichloromethane (10 ml) was added dropwise. After stirring at 0°C for 60 minutes, the reaction was poured into cold water (20 ml). The organic phase was separated and washed with a cold 5% aqueous solution of NaOH, water, dried (Na<sub>2</sub>SO<sub>4</sub>), and the solvent removed in vacuo to leave a yellow-brown oil.  
20 Purification by Kugelrohr distillation gave the product as a clear oil (0.88 g, 76%). <sup>1</sup>H NMR (300MHz, CDCl<sub>3</sub>)  $\delta$  4.12 (s, 2H, -CH<sub>2</sub>-), 7.44 (s, 5H, arom). <sup>13</sup>C NMR (100MHz, CDCl<sub>3</sub>)  $\delta$  33.2 (-CH<sub>2</sub>-), 129.3 (arom), 129.8 (arom), 134.9 (arom), 190.7 (-C=O). MS (PCI, NH<sub>3</sub>) 232 [M+ H]<sup>+</sup>.

The modifier shown in Figure 3 has been derived from 1,2-amiothiol  
25 reactive groups described for orthogonal ligation of peptide fragments (Dawson et al., Science 266, 776-779 (1994); Liu & Tam Proc. Natl. Acad. Sci. USA 91, 6584-6588 (1994)).

Alkylation of 5'-thiophosphate RNA with phenyl- $\alpha$ -bromothioacetate (Figure 3) is illustrated in Figure 4. This alkylation step has been carried out as follows. 10  $\mu$ M GMPS-RNA labeled with  $^{32}$ P was reacted with 8 mM phenyl-bromothioacetate in 8% DMSO, 82 mM sodium phosphate buffer, pH6.8, at room temperature for 40 minutes. After reaction, the mixture was extracted 4 times with chloroform to remove unreacted bromide. Precipitation was avoided because of the possibility of exchanging the thioester with ethanol.

#### Conjugate Formation Using Orthogonal Ligation

Orthogonal ligation of protein and nucleic acid according to this first method is based on a fast chemoselective thiol-exchange followed by intramolecular amide bond formation, leading to a covalent link between a nucleic acid and a protein. This method, which is illustrated diagrammatically in Figure 5, allows efficient ligation of RNA and peptide at  $\mu$ M concentrations of reactants. When this reaction has been carried out, no side products have been detected.

In one particular ligation reaction, 2.5  $\mu$ M thioester RNA of the following sequence (SEQ ID NO: 1):

thiophosphate-GGG-N80-CCGUGAAGAGCAUUGG

was reacted with 25  $\mu$ M peptide 1 (CSKGFGFVSFSYK-biotin; SEQ ID NO: 2), 25  $\mu$ M peptide 2 (CRKKRRQRRRPPQGSQTHQVLSKQK-biotin; SEQ ID NO: 3), or 25  $\mu$ M peptide 3 (MSKGFGFVSFSYK-biotin; SEQ ID NO: 4) in 80 mM sodium phosphate buffer pH6.8 and 0.5% thiophenol for 2 hours at 30°C. After reaction, the RNA was purified on a polyacrylamide gel and then bound to neutravidin-agarose (Pierce). Bound RNA was eluted with 10  $\mu$ g/ml proteinase K for 5 minutes. Scintillation counting revealed that 10-12% of the RNA was linked to biotinylated peptides 1 and 2 carrying an N-terminal

cysteine, whereas peptide 3 reacted with less than 0.2% of the RNA.

In a further experiment, 1  $\mu$ M thioester-RNA was reacted with 1 mM peptide 2 under the conditions described above, for 3 hours or 20 hours. The reactions were analyzed by electrophoresis using a 6% polyacrylamide  
5 TBE/urea gel (Novex). Under these conditions, 50% of the RNA had reacted in less than 3 hours, but no additional reaction was observed following a prolonged incubation.

Orthogonal ligation may also be used to ligate RNA and protein while these complexes are bound to the ribosome, either during or after  
10 translation (see Figure 2), thereby generating 5'-fusions between an mRNA and its encoded peptide in a pseudo-intermolecular reaction. In one preferred method, the mRNA is used in a cell-free translation system and shows the following properties: (1) the mRNA carries a 1,2-aminothiol reactive group at its 5'-end; (2) the mRNA encodes an N-terminal protease recognition sequence  
15 followed by the amino acid cysteine; (3) the mRNA codes for a protein which is at least 40-50 amino acids long; and (4) the mRNA is devoid of stop codons.

The defined minimal protein length of 40-50 amino acids ensures that the N-terminus of a nascent protein extends to the surface of the ribosome, thus exposing the recognition sequence to protease cleavage. The absence of  
20 stop codons prevents release of the mRNA from the ribosome. Addition of Mg salt and washing buffer at low temperature stalls and stabilizes the mRNA-ribosome-protein complex after translation (Hanes & Plueckthun, Proc. Natl. Acad. Sci. USA 94, 4937-4942 (1997)). Protease treatment may be carried out in this same buffer to expose the N-terminal cysteine on the nascent,  
25 ribosome-bound protein. Subsequently, orthogonal ligation between the 5'-terminal 1,2-aminothiol reactive group and the N-terminal cysteine can take place, leading to fusions between nascent proteins and their encoding mRNAs.

To further enhance the ability to efficiently form fusions on the ribosome, stalled mRNA-ribosome-protein complexes (prepared, for example, by the method of Hanes & Plueckthun, Proc. Natl. Acad. Sci. USA 94, 4937-4942 (1997)) may be prepared from cell-free translation systems in which the concentration of cysteine is reduced. Preparation of lysates which are devoid or which contain only a minimal amount of cysteine (preferably, < 1  $\mu$ M) have been described (see, for example, the instruction manual on in vitro translation kits, Ambion, TX). A low concentration of competing free cysteine in the lysate may increase the efficiency of productive orthogonal ligation reactions between the N-terminal cysteine of an encoded protein and the 5'-terminal 1,2 aminothiols reactive group, thus increasing RNA-protein fusion yields.

#### Bisarsenical-Tetracysteine Conjugate Formation

An alternative method for the conjugation of nucleic acids and proteins is through a bisarsenical-tetracysteine interaction. This method of conjugate formation relies on the affinity of organic arsenicals for sulfhydryl-containing compounds (Webb, in Webb (ed.), Enzyme and Metabolic Inhibitors, vol. 3, Academic Press, New York 1966, Cullen et al., J. Inorg. Biochem 21, 179 (1984)), an interaction which has been utilized successfully in the in vivo, sequence-specific identification of fusion proteins which carry non-native sequences consisting of tetracysteine motifs within  $\alpha$ -helical structures (Griffin et al., Science 281, 269-272 (1998)). The technique is shown schematically in Figure 6.

As shown in Figure 6, the 5'-terminus of the mRNA is modified with a bisarsenical derivative which is capable of binding an  $\alpha$ -helical tetracysteine motif. The modified message encodes an amino acid sequence which is chosen

for, or designed to have, a propensity to form  $\alpha$ -helices under physiological conditions. Such a modified message may contain a nucleic acid sequence that encodes an amino acid sequence chosen for its propensity to form  $\alpha$ -helices under conditions compatible with in vitro translation. A tetracysteine motif of the form CysCysXXCysCys is included within the helix to create the necessary geometry for thiol exchange. The cys4  $\alpha$ -helix is formed preferably at the N-terminus of the encoded protein. This motif may either be introduced through mutation of an existing  $\alpha$ -helix within the native protein (for example, by the approach of Griffin et al., Science 281, 269-272 (1998)) or by fusion of the motif to the N-terminus of the protein of interest (for example, during chemical protein synthesis). A tetracysteine motif of the form, cys, cys+1, cys+4, cys+5 is included within the helix to create the necessary geometry for bisarsenical chelation. A tricyclic scaffold is used to allow sufficient spatial orientation of the dithiarsolane moieties to bind the tetracysteine motif effectively. The bisarsenical derivative features a reactive moiety for the regiospecific attachment of the compound to the nucleic acid terminus. This attachment functionality may also be used for derivatization of the bisarsenical compound to a solid phase.

One exemplary scheme for the synthesis of a bisarsenical derivative which encompasses the above features is outlined in Figure 7. The tricyclic scaffold, 4,5-diiodo-9(10H)-anthracenone **4** is constructed from 1,8-dichloroanthraquinone **1** using standard methods (as described, for example, in Lovell & Joule; Synth. Commun. 27(7), 1209-1215 (1997)). The anthracenone nucleus serves as a handle to introduce a linker via O-alkylation to form compound **5**, as described, for example, in Johnstone and Rose (Tetrahedron 35, 2169-2173 (1979)) or Loupy et al. (Bull. Soc. Chim. Fr. 1027-1035 (1987)). Dithiarsolane formation may be achieved by transmetallation via

transition metal-mediated catalysis (as described, for example, in Griffin et al., Science 281, 269-272 (1998)) with concomitant reaction with the appropriate dithiol. Introduction of the attachment moiety via carboxylic acid-activated amide formation completes the synthesis of 7. This step may be carried out as  
5 described, for example, in Desai and Stramiello, Tet. Letts. 34 (48), 7685-7688 (1993).

Another scheme for preparing an amino-tethered bisarsenical fluorescein derivatives is described by Thorn et al., Protein Science 9: 213-217 (2000). Reaction with succinimidyl 4-(p-maleimidophenyl) butyrate (SMPB,  
10 Pierce, Rockford, IL) yields a maleic imid-tethered derivative of bisarsenical fluorescein (as shown in Figure 8).

These tethered derivatives (compound 7 in Figure 7) and (compound 9 in Figure 8) may be attached to the 5' end of a 5' thiol RNA, for example, by the method of Hermanson, Bioconjugate Techniques, Academic Press, San  
15 Diego CA (1996); and Goodchild in Meares (ed.), Perspectives in Bioconjugate Chemistry, American Chemical Society, Washington, DC 1993. This putative cys4-helix binding molecule may also mediate the formation of nucleic-acid protein conjugates through attachment at the 3'-terminus of the nucleic acid (Cremer et al., J. Protein Chem. 11(5), 553-560 (1992). The conjugation  
20 reaction between the nucleic acid carrying the bisarsenical derivative and the protein may be carried out in buffer or lysate.

Other embodiments are within the claims.

What is claimed is:

Claims

1. A method for generating a 5'-nucleic acid-protein conjugate, said method comprising:
- 5       (a) providing a nucleic acid which carries a reactive group at its 5' end;
- (b) providing a non-derivatized protein; and
- (c) contacting said nucleic acid and said protein under conditions which allow said reactive group to react with the N-terminus of said protein, thereby forming a 5'-nucleic acid-protein conjugate.
- 10       2. The method of claim 1, wherein said nucleic acid is greater than about 20 nucleotides in length; greater than about 120 nucleotides in length; or between about 2-1000 nucleotides in length.
3. The method of claim 1, wherein said protein is greater than about 20 amino acids in length; greater than about 40 amino acids in length; or
- 15       between about 2-300 amino acids in length.
4. The method of claim 1, wherein said contacting step is carried out in a physiological buffer.
5. The method of claim 1, wherein said contacting step is carried out using a nucleic acid and a protein, both of which are present at a concentration
- 20       of less than about 1 mM.

15. A 5'-nucleic acid-protein conjugate produced by the method of claim 1.

16. A 5'-nucleic acid-protein conjugate comprising a nucleic acid bound through its 5'-terminus or a 5'-terminal reactive group to the N-terminus  
5 of a non-derivatized protein.

17. The conjugate of claim 16, wherein said conjugate is immobilized on a solid support.

18. The conjugate of claim 17, wherein said solid support is a bead or chip.

10 19. The conjugate of claim 17, wherein said conjugate is one of an array immobilized on said solid support.

20. The conjugate of claim 16, wherein said nucleic acid is greater than about 20 nucleotides in length.

21. The conjugate of claim 16, wherein said protein is greater than  
15 about 20 amino acids in length.

22. The conjugate of claim 16, wherein said nucleic acid is DNA or RNA.

23. The conjugate of claim 16, wherein said nucleic acid comprises the coding sequence for said protein.

24. The conjugate of claim 16, wherein said N-terminus of said non-derivatized protein is a cysteine residue.

25. The conjugate of claim 16, wherein said protein comprises an  $\alpha$ -helical tetracysteine motif located proximal to its N-terminus.

5           26. The conjugate of claim 25, wherein said  $\alpha$ -helical tetracysteine motif comprises cys-cys-X-X-cys-cys, wherein X is any amino acid.

27. A method for the selection of a desired nucleic acid or a desired protein, said method comprising:

10           (a) providing a population of 5'-nucleic acid-protein conjugates, each comprising a nucleic acid bound through its 5'-terminus or a 5'-terminal reactive group to the N-terminus of a non-derivatized protein;

            (b) contacting said population of 5'-nucleic acid-protein conjugates with a binding partner specific for either the nucleic acid or the protein portion of said desired nucleic acid or desired protein under conditions which allow for  
15 the formation of a binding partner-candidate conjugate complex; and

            (c) substantially separating said binding partner-candidate conjugate complex from unbound members of said population, thereby selecting said  
desired nucleic acid or said desired protein.

28. The method of claim 27, wherein said method further comprises  
20 repeating steps (b) and (c).

6. The method of claim 1, wherein said nucleic acid is DNA or RNA.
7. The method of claim 6, wherein said RNA is mRNA.
8. The method of claim 1, wherein said nucleic acid comprises the  
5 coding sequence for said protein.
9. The method of claim 1, wherein said N-terminus of said non-derivatized protein is a cysteine residue.
10. The method of claim 9, wherein said N-terminal cysteine is exposed by protein cleavage.
- 10 11. The method of claim 9, wherein said reactive group is an aminothioli reactive group.
12. The method of claim 1, wherein said protein comprises an  $\alpha$ -helical tetracysteine motif located proximal to its N-terminus.
13. The method of claim 12, wherein said  $\alpha$ -helical tetracysteine  
15 motif comprises cys-cys-X-X-cys-cys, wherein X is any amino acid.
14. The method of claim 12, wherein said reactive group is a bisarsenical derivative.

29. A method for detecting an interaction between a protein and a compound, said method comprising:

(a) providing a solid support comprising an array of immobilized 5'-nucleic acid-protein conjugates, each conjugate comprising a nucleic acid  
5 bound through its 5'-terminus or a 5'-terminal reactive group to the N-terminus of a non-derivatized protein;

(b) contacting said solid support with a candidate compound under conditions which allow an interaction between said protein portion of said conjugate and said compound; and

10 (c) analyzing said solid support for the presence of said compound as an indication of an interaction between said protein and said compound.

30. The method of claim 29, wherein said solid support is a bead or a chip.

31. The method of claim 29, wherein said compound is a protein.

15 32. The method of claim 35, wherein said compound is a therapeutic.

1/8

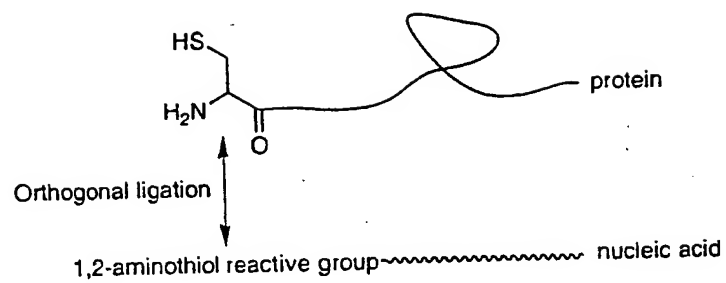


Figure 1

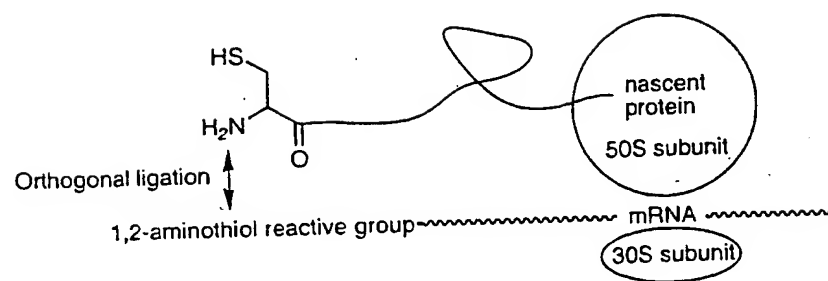


Figure 2

3/8

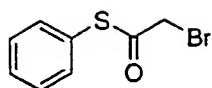


Figure 3

4/8

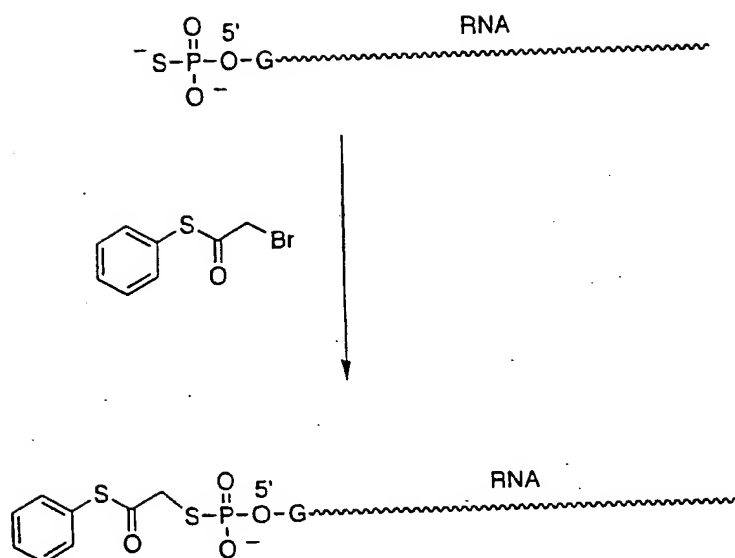


Figure 4

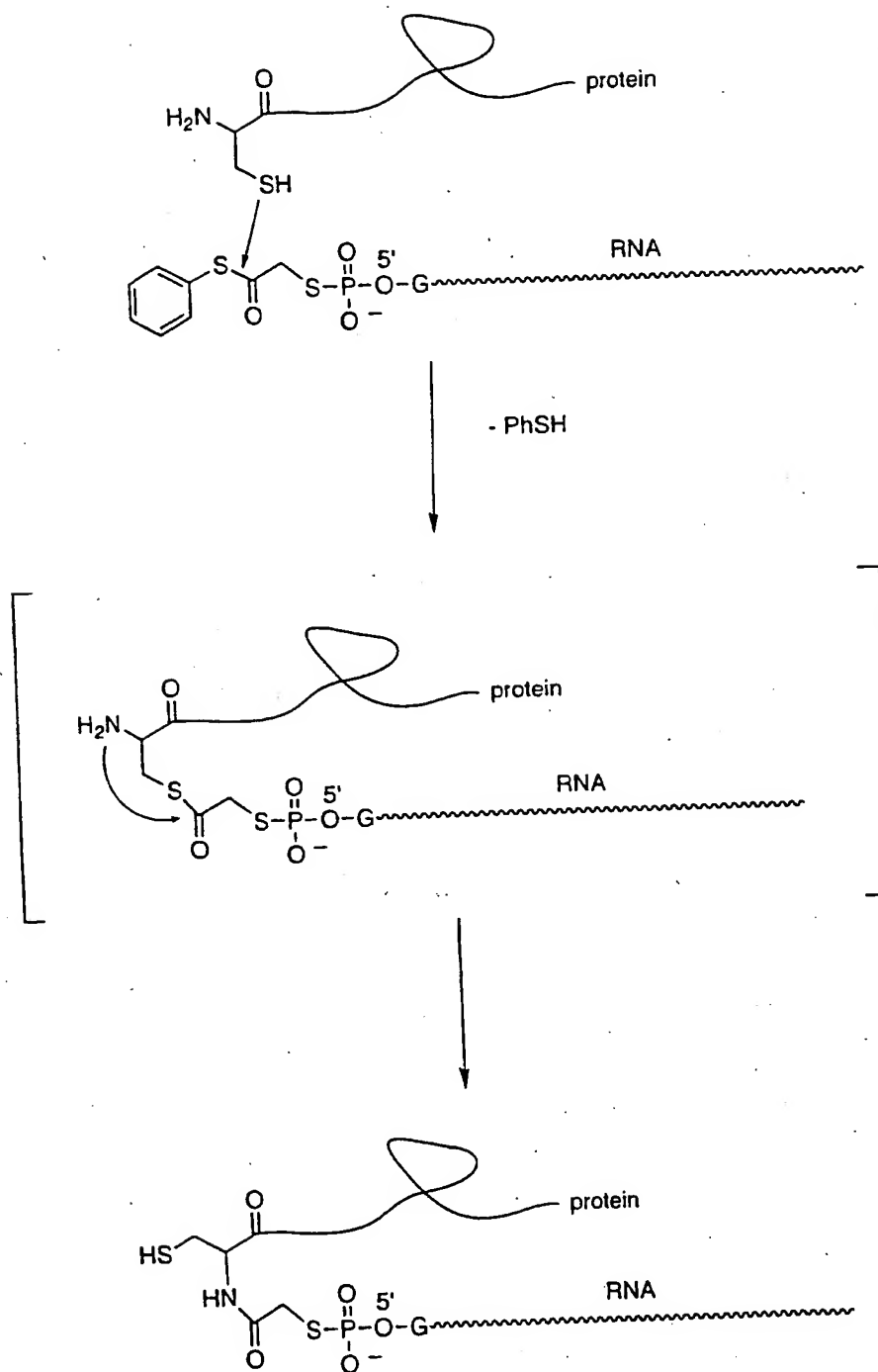


Figure 5

6/8

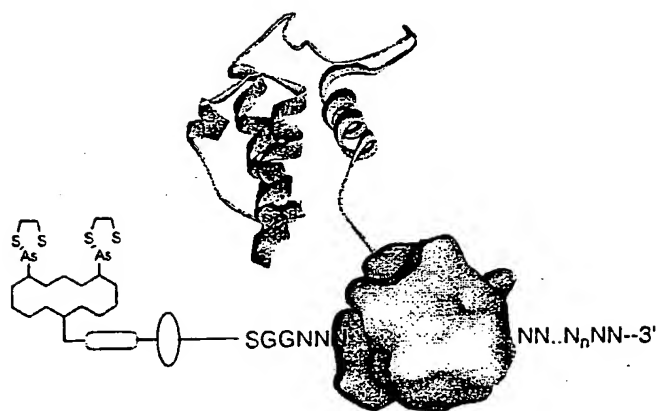


Figure 6

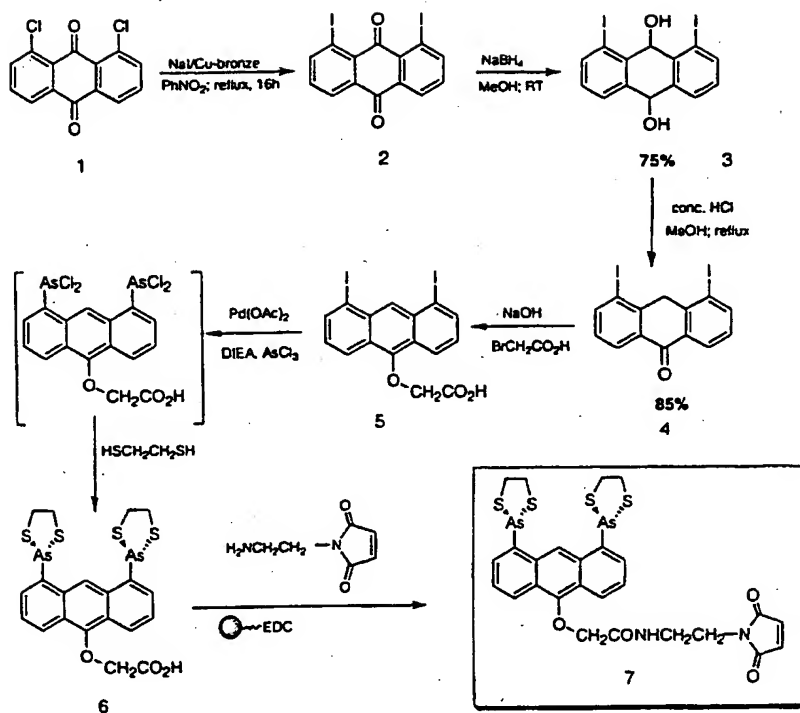


Figure 7

8/8

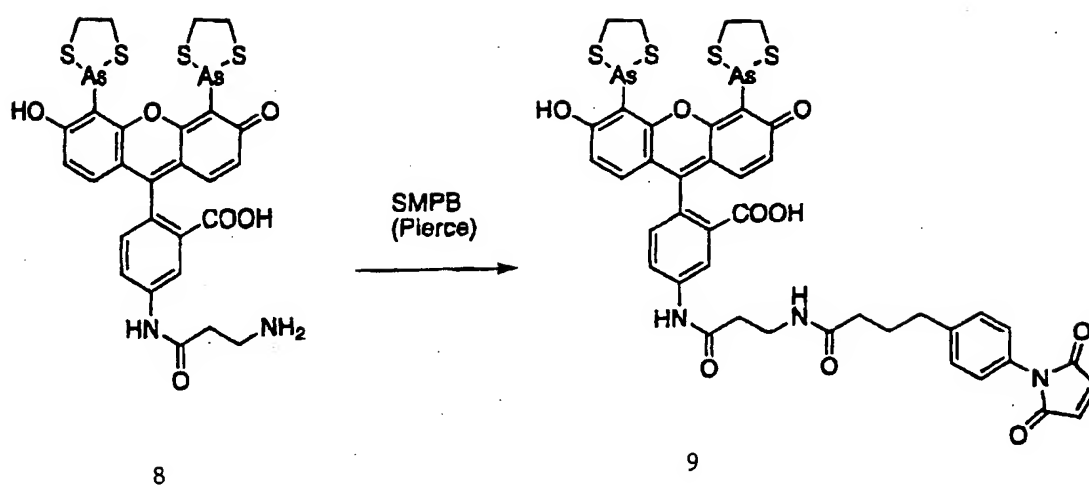


Figure 8

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US00/15077

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
IPC(7) : A61K 38/16, 38/03; C07K 14/00		
US CL : 530/300, 350; 514/2, 12, 13, 14, 15, 16, 17, 18, 19		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols)		
U.S. : 530/300, 350; 514/2, 12, 13, 14, 15, 16, 17, 18, 19		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
MEDLINE, BIOSIS, CAPLUS, EMBASE, LIFESCI, WEST		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 263 740 A1 (LEBLEU et al) 13 April 1988, Abstract.	1-32
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *A* document defining the general state of the art which is not considered to be of particular relevance *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *E* earlier document published on or after the international filing date *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *G* document member of the same patent family	
Date of the actual completion of the international search		Date of mailing of the international search report
20 JULY 2000		17 AUG 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231		Authorized officer
Facsimile No. (703) 305-3230		STEPHEN TU
		Telephone No. (703) 308-0196